A frailty model for (interval) censored family survival data, applied to the age at onset of non-physical problems

Jonker, M.A.; Boomsma, D.I.

published in
Lifetime data analysis
2010

DOI (link to publisher)
10.1007/s10985-009-9141-5

document version
Publisher's PDF, also known as Version of record

Lifetime Data Anal (2010) 16:299–315
DOI 10.1007/s10985-009-9141-5

A frailty model for (interval) censored family survival data, applied to the age at onset of non-physical problems

M. A. Jonker · D. I. Boomsma

Received: 27 March 2009 / Accepted: 6 November 2009 / Published online: 25 November 2009

Abstract Family survival data can be used to estimate the degree of genetic and environmental contributions to the age at onset of a disease or of a specific event in life. The data can be modeled with a correlated frailty model in which the frailty variable accounts for the degree of kinship within the family. The heritability (degree of heredity) of the age at a specific event in life (or the onset of a disease) is usually defined as the proportion of variance of the survival age that is associated with genetic effects. If the survival age is (interval) censored, heritability as usually defined cannot be estimated. Instead, it is defined as the proportion of variance of the frailty associated with genetic effects. In this paper we describe a correlated frailty model to estimate the heritability and the degree of environmental effects on the age at which individuals contact a social worker for the first time and to test whether there is a difference between the survival functions of this age for twins and non-twins.

Keywords Additive gamma frailty model · Twin study · Family data · Survival analysis · Interval censoring · Heritability

1 Introduction

Family survival data are often used in genetic research to estimate genetic and environmental contributions to age at onset of a disease or of a specific event in life, as the

M. A. Jonker (✉)
Department of Mathematics, Faculty of Sciences, VU University Amsterdam,
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
e-mail: majonker@few.vu.nl

D. I. Boomsma
Department of Biological Psychology, VU University, Amsterdam, The Netherlands

age at which individuals need social help for the first time (like in this paper). The data can be modeled with a correlated frailty model (see, e.g., Yashin and Iachine 1999; Iachine 2001; Jonker et al. 2009).

Frailty models were originally introduced to model missing covariates. Later, Vaupel et al. (1979) extended these models to include single shared frailties to model simple patterns of dependence between the survival times of related individuals. In the single shared frailty model all individuals in a group share the same frailty, but the frailties of individuals from different groups are independent. In genetic research a group of individuals usually consists of individuals from the same family; individuals who share a part of their genes. The shared frailty model is too simplistic for these kind of data, because different family bands should correspond with different frailty correlations. To deal with more complex dependence structures between individuals, the shared frailty models were extended to correlated frailty models (Yashin et al. 1995; Andersen et al. 1992).

Our aim is to estimate the degree of heredity (heritability), environmental effects and twin effects of the age at which people contact a social helper for the first time, to test whether these terms differ for males and females, and to investigate whether the survival functions differ for twins and their non-twin siblings. Our data come from an ongoing study on health, lifestyle, and personality. Longitudinal data were collected of Dutch monozygotic and dizygotic twins and of their siblings. At five different time points between 1991 and 2002 the twins and their siblings were asked whether they ever contacted a social worker for non-physical problems. So, not the age at which an individual contacted a social worker for the first time is observed, but an age interval in which this age falls. Summarized, we have family survival data which consists of interval censored age at onset data for monozygotic or dizygotic twins and of their siblings. We model the data with a correlated gamma frailty model. The frailty term of each individual is decomposed as a linear combination of four independent gamma distributed random variables. These terms represent the genetic contribution to the age at onset, contributions by the common environment of all siblings (twins and non-twins), a twin-effect and the last term represents the contribution by individual specific, unshared alleles and environment. We include a twin-effect in the model, since environmental/social effects may differ for twins and non-twins.

Heritability is usually defined as the proportion of variance of the quantitative trait or survival age that is associated with genetic effects. Since the survival age is never observed, heritability as usually defined can not be estimated. Instead, we define heritability as the proportion of variance of the frailty associated with genetic effects. So, the definition of heritability is as usual, but the frailties are viewed as the latent phenotypes (see, e.g., Yashin and Iachine 1999; Iachine 2001; Jonker et al. 2009). The same arguments hold for the environmental and twin effects.

For mathematical convenience we assume that the frailty variable follows a gamma distribution. In that case, an explicit expression of the simultaneous survival function for the twins and their siblings can be expressed in terms of the marginal survival function and the other model parameters. All unknown parameters are estimated by their maximum likelihood estimators and statistical tests are performed with a likelihood ratio test.

The remainder of the paper is organized as follows. In Sect. 2 we describe the additive gamma frailty model for family survival data. Next, in Sect. 3, we give an expression of the likelihood function for the interval censored data as in our data-set. This likelihood is maximized to obtain the maximum likelihood estimators of the unknown parameters in the model. We furthermore describe statistical tests for testing whether genetic, environmental, and twin effects are significantly present, whether these terms differ for males and females and whether the age at onset distribution of twins and of their non-twin siblings differ. The results of the analysis is described in Sect. 4 and we conclude the paper with a discussion in Sect. 5. Mathematical details are given in the Appendix.

2 The model

Suppose that survival data of n families is available. Each family consists of a twin (monozygotic or dizygotic) and at most two siblings (if data of more than two siblings is available, a part of the data is simply ignored). We assume that data from different families are independent. For ease of notation, we describe the model for only one foursome (a twin and two siblings); for a couple or a trio the model directly follows from the model for the foursomes. Let (T_1, T_2, T_3, T_4) be the survival ages (the age at which the individuals contacted a social worker for the first time) of the four individuals of a family/a foursome. The first and the second coordinates always refer to the twin (the order of the twin halves is of no importance), the third and fourth coordinates are the survival ages of the siblings. Furthermore, let (Z_1, Z_2, Z_3, Z_4) be a foursome of latent variables ("frailties"), such that T_1, T_2, T_3 and T_4 are conditionally independent given the frailties (Z_1, Z_2, Z_3, Z_4) with hazard functions $t \to Z_1\lambda(t), t \to Z_2\lambda(t), t \to Z_3\lambda(t)$ and $t \to Z_4\lambda(t)$, respectively, for a given "baseline hazard function" λ. For every individual in the foursome, we decompose the frailty variable into a sum of four independent components:

$$Z_i = A_i + C + B_i + E_i, \quad \text{for } i = 1, 2, 3, 4.$$

The first component, A_i, represents the additive genetics for individual i, the second component C is for the common environment of the four siblings (twin halves and siblings), and thus is equal for all four siblings, the third component reflects the specific twin environmental effects, and the fourth component, E_i, represents the non-shared, specific environmental and genetic effects for the ith individual in the foursome. We assume that the vectors (A_1, A_2, A_3, A_4) and (B_1, B_2, B_3, B_4) and the variables C, and E_1, E_2, E_3 and E_4 are independent. We moreover assume that the correlation between the genetic effects of two siblings equals 1/2 unless the siblings form a monozygotic twin; then the correlation between the two terms equals 1 ($\text{Cor}(A_i, A_j) = 1/2$ for $(i \neq j)$, unless individuals i and j form a monozygotic twin, then $\text{Cor}(A_i, A_j) = 1$). We furthermore assume that the correlation between the twin-effect terms equals 1 for twins and 0 for non-twins ($\text{Cor}(B_1, B_2) = 1, \text{Cor}(B_1, B_3) = \text{Cor}(B_1, B_4) = \text{Cor}(B_2, B_3) = \text{Cor}(B_2, B_4) = \text{Cor}(B_3, B_4) = 0$). The terms $E_1, \ldots, E_4$ are assumed to be mutually independent. Furthermore, we assume that all terms in the frailty

decomposition have gamma distributions with inverse scale parameter η and shape parameters ν for the terms A_1, A_2, A_3 and A_4, ν_c for C, ν_b for B_1, B_2, B_3 and B_4, and shape parameter ν_e for E_1, E_2, E_3 and E_4. We set $\eta = \nu + \nu_c + \nu_b + \nu_e$ so that $\mathrm{E}Z_1 = \mathrm{E}Z_2 = \mathrm{E}Z_3 = \mathrm{E}Z_4 = 1$ (see also Yashin et al. 1999; Jonker et al. 2009).

The proportions of variance of individual frailties associated with additive genetic effects (h^2), shared environmental effects (c^2), twin social effects (b^2) and non-shared environmental and genetic effects (e^2) are defined as

$$h^2 := \frac{\nu}{\eta}, \quad c^2 := \frac{\nu_c}{\eta}, \quad b^2 := \frac{\nu_d}{\eta} \quad \text{and} \quad e^2 := \frac{\nu_e}{\eta}.$$

Then $h^2 + c^2 + b^2 + e^2 = 1$, since $\eta = \nu + \nu_c + \nu_b + \nu_e$, and the correlation between the frailties of monozygotic and dizygotic twin halves and siblings equal:

$$\begin{aligned}
\rho_{MZ} &= \frac{\nu}{\eta} + \frac{\nu_c}{\eta} + \frac{\nu_b}{\eta} \;=\; h^2 + c^2 + b^2, \\
\rho_{DZ} &= \frac{1}{2}\frac{\nu}{\eta} + \frac{\nu_c}{\eta} + \frac{\nu_b}{\eta} \;=\; \frac{1}{2}h^2 + c^2 + b^2, \\
\rho_{S,S} &= \frac{1}{2}\frac{\nu}{\eta} + \frac{\nu_c}{\eta} \;=\; \frac{1}{2}h^2 + c^2,
\end{aligned} \tag{1}$$

with $\rho_{S,S}$ the correlation between the frailties of two siblings who do not form a twin (see Appendix for the computations).

Define $S_{MZ,S,S}$ and $S_{DZ,S,S}$ as the simultaneous survival functions for a foursome that consists of a monozygotic twin and two siblings, and a dizygotic twin and two siblings, respectively. The vector of frailties is assumed to be multivariate gamma distributed. In this special case the simultaneous survival functions $S_{MZ,S,S}$ and $S_{DZ,S,S}$ can be written in terms of the marginal survival function S (see the Appendix for the derivation):

$$\begin{aligned}
&S_{MZ,S,S}(t_1,t_2,t_3,t_4) = \mathrm{P}(T_1 > t_1, T_2 > t_2, T_3 > t_3, T_4 > t_4 | MZ \times Sib \times Sib) \\
&= \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 3}\right)^{\nu/4+\nu_c} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 2}\right)^{\nu/4} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 2}\right)^{\nu/4} \left(\frac{1}{S(t_3)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 1}\right)^{\nu/4} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1}\right)^{\nu/4+\nu_b} (S(t_1)S(t_2))^{\nu_e\sigma^2} (S(t_3)S(t_4))^{(\nu/4+\nu_b+\nu_e)\sigma^2}.
\end{aligned} \tag{2}$$

and

$$S_{DZ,S,S}(t_1,t_2,t_3,t_4) = \mathrm{P}(T_1 > t_1, T_2 > t_2, T_3 > t_3, T_4 > t_4 | DZ \times Sib \times Sib)$$

$$
\begin{aligned}
&= \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 3}\right)^{\nu/8+\nu_c} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 2}\right)^{\nu/8} \\
&\quad \times \left(\frac{1}{S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 2}\right)^{\nu/8} \\
&\quad \times \left(\frac{1}{S(t_3)^{-\sigma^2} + S(t_4)^{-\sigma^2} + S(t_1)^{-\sigma^2} - 2}\right)^{\nu/8} \\
&\quad \times \left(\frac{1}{S(t_4)^{-\sigma^2} + S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 2}\right)^{\nu/8} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1}\right)^{\nu/8+\nu_b} \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 1}\right)^{\nu/8} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 1}\right)^{\nu/8} \left(\frac{1}{S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 1}\right)^{\nu/8} \\
&\quad \times \left(\frac{1}{S(t_2)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 1}\right)^{\nu/8} \left(\frac{1}{S(t_3)^{-\sigma^2} + S(t_4)^{-\sigma^2} - 1}\right)^{\nu/8} \\
&\quad \times (S(t_1)S(t_2))^{(\nu/8+\nu_e)\sigma^2} (S(t_3)S(t_4))^{(\nu/8+\nu_b+\nu_e)\sigma^2}. \qquad (3)
\end{aligned}
$$

with $\sigma^2 = 1/\eta$ the variance of Z_1, Z_2, Z_3 and Z_4. Note that $S_{MZ,S,S}(t,0,0,0) = S_{MZ,S,S}(0,t,0,0) = S_{MZ,S,S}(0,0,t,0) = S_{MZ,S,S}(0,0,0,t) = S(t)$ is independent of the relationship between the four individuals. The same holds for $S_{DZ,S,S}$.

If a family consists of a twin and one sibling, an expression for their simultaneous survival function in terms of the marginal survival function can be found by setting t_4 in $S_{MZ,S,S}(t_1,t_2,t_3,t_4)$ and $S_{DZ,S,S}(t_1,t_2,t_3,t_4)$ equal to zero:

$$
\begin{aligned}
S_{MZ,S}(t_1,t_2,t_3) &= \mathrm{P}(T_1 > t_1, T_2 > t_2, T_3 > t_3 | MZ \times Sib) \\
&= \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 2}\right)^{\nu/2+\nu_c} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1}\right)^{\nu/2+\nu_b} \\
&\quad \times S(t_3)^{(\nu/2+\nu_b)\sigma^2} (S(t_1)S(t_2)S(t_3))^{\nu_e\sigma^2} \qquad (4)
\end{aligned}
$$

and

$$
\begin{aligned}
S_{DZ,S}(t_1,t_2,t_3) &= \mathrm{P}(T_1 > t_1, T_2 > t_2, T_3 > t_3 | DZ \times Sib) \\
&= \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 2}\right)^{\nu/4+\nu_c} \\
&\quad \times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1}\right)^{\nu/4+\nu_b}
\end{aligned}
$$

$$\times \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 1}\right)^{\nu/4} \left(\frac{1}{S(t_2)^{-\sigma^2} + S(t_3)^{-\sigma^2} - 1}\right)^{\nu/4}$$
$$\times (S(t_1)S(t_2)S(t_3))^{(\nu/4+\nu_e)\sigma^2} S(t_3)^{\nu_b\sigma^2}. \quad (5)$$

If the value at another coordinate is set equal to zero, the survival function of another trio is found.

Next, we set $t_3 = 0$ in $S_{MZ,S}(t_1, t_2, t_3)$ and $S_{DZ,S}(t_1, t_2, t_3)$ to find expressions for the simultaneous survival function for a monozygotic twin, S_{MZ}, and a dizygotic twin, S_{DZ}:

$$S_{MZ}(t_1, t_2) = \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1}\right)^{\rho_{MZ}/\sigma^2} S(t_1)^{1-\rho_{MZ}} S(t_2)^{1-\rho_{MZ}}, \quad (6)$$

for monozygotic twins and

$$S_{DZ}(t_1, t_2) = \left(\frac{1}{S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1}\right)^{\rho_{DZ}/\sigma^2} S(t_1)^{1-\rho_{DZ}} S(t_2)^{1-\rho_{DZ}}, \quad (7)$$

for dizygotic twins with ρ_{DZ} and ρ_{MZ} as defined before.

The marginal survival function S can be left completely unknown, or we may assume that it belongs to a family of parametric survival functions which are known up to a finite dimensional parameter.

2.1 Gender effects

If the marginal survival functions for males and females differ, S is replaced by gender specific survival functions. Also if the survival functions of twins and of their non-twin siblings are thought to be different, twin and non-twin specific survival functions can be used instead.

Suppose one is interested in the question whether the degree of heritability differs between males and females. Then, the frailty decomposition in the model can be extended with an extra term:

$$Z_i^f = A_i + A_i^f + C + B_i + E_i^f,$$

for females, and

$$Z_i^m = A_i + A_i^m + C + B_i + E_i^m,$$

for males. In these decompositions A_i reflects the genetic additive effects that males and females have in common. The gender specific terms, A_i^f and A_i^m, represent the genetic effects that are specific for females and males. So, the genetic additive effect for females is given by $A_i + A_i^f$ and for males by $A_i + A_i^m$. The distributions and the mutual correlations (also among different family members) of A_i, C and B_i are defined

as they were defined in the non-gender specific model in the previous subsection. We assume that A_i^f and A_j^m $(i \neq j)$ are mutually independent and are independent of A_k, C, B_k and E_k^m and E_k^f for $i, j, k = 1, 2, 3, 4$. The correlation $\mathrm{Cor}(A_i^f, A_j^f) = 1/2$ for $(i \neq j)$ unless $i = 1$, $j = 2$ and the first and second individual in the family form a monozygotic-twin, then $\mathrm{Cor}(A_1^f, A_2^f) = 1$. The same correlations hold for the additive genetic male terms. If the individuals are of opposite sex, the correlation equals zero: $\mathrm{Cor}(A_i^f, A_j^m) = 0$ for $i, j = 1, 2, 3, 4$, $i \neq j$. We furthermore assume that A_i^f and A_j^m have gamma distributions with shape parameters ν^f and ν^m and inverse scale parameter η. The variables E_i^f and E_j^m represent the non-shared, specific, environmental and genetic effects for females and males. The distributions of these terms are gender specific, since the genetic additive effects between males and females may differ. We assume that E_i^f and E_j^m are independent of each other and are independent of all other terms in the frailty decomposition. E_i^f and E_j^m have a gamma distribution with inverse scale parameter η and with shape parameters ν_e^f and ν_e^m, respectively. We assume that $\nu^f + \nu_e^f = \nu^m + \nu_e^m$, then $\nu + \nu^f + \nu_c + \nu_b + \nu_e^f = \nu + \nu^m + \nu_c + \nu_b + \nu_e^m$. We furthermore assume that $\eta = \nu + \nu^f + \nu_c + \nu_b + \nu_e^f = \nu + \nu^m + \nu_c + \nu_b + \nu_e^m$, so that $\mathrm{E}Z_i^f = \mathrm{E}Z_i^m = 1$ for $i = 1, \ldots, 4$. Define $h_f^2 = \nu^f/\eta$ and $h_m^2 = \nu^m/\eta$ as the gender specific heritabilities for females and males. Then, the heritability for females is given by $h^2 + h_f^2$ and for males by $h^2 + h_m^2$. In Sect. 3 a statistical test is described for testing whether h^2, h_f^2 and h_m^2 are positive.

If one is interested in the question whether the common environment or the twin effect is gender specific, the frailty model can be extended in a similar way.

3 Estimation and testing

3.1 Estimation

The survival ages of the twin halves and their siblings T_1, T_2, T_3 and T_4 are not observed. In stead intervals $[U_1, V_1], [U_2, V_2], [U_3, V_3]$ and $[U_4, V_4]$ are observed with $U_i \leq T_i \leq V_i$ for $i = 1, 2, 3, 4$. Possibly $U_i = 0$ and/or $V_i = \infty$. If $U_i = 0$ and $V_i = v$, individual i had contacted a social worker before age v. However, if $U_i = u$ and $V_i = \infty$, individual i did not contact a social worker before the age of u, and if $U_i = 0$ and $V_i = \infty$ no information for individual i is available. It is reasonable to assume that the distribution functions of the variables U_i and V_i do not depend on the frailty parameters and the parameters for the marginal survival functions. When maximizing the likelihood with respect to all parameters of interest, the terms in the likelihood with only the distribution functions of the variables U_i and V_i can be left out. Then, the likelihood for a foursome consisting of a monozygotic twin and two siblings is proportional to

$$\begin{aligned}&\mathrm{lik}_{MZ,S,S}((U_1, V_1), (U_2, V_2), (U_3, V_3), (U_4, V_4))\\ &= S_{MZ,S,S}(U_1, U_2, U_3, U_4) - S_{MZ,S,S}(U_1, U_2, U_3, V_4) - S_{MZ,S,S}(U_1, U_2, V_3, U_4)\end{aligned}$$

$$
\begin{aligned}
&- S_{MZ,S,S}(U_1, V_2, U_3, U_4) - S_{MZ,S,S}(V_1, U_2, U_3, U_4) + S_{MZ,S,S}(U_1, U_2, V_3, V_4)\\
&+ S_{MZ,S,S}(U_1, V_2, U_3, V_4) + S_{MZ,S,S}(V_1, U_2, U_3, V_4) + S_{MZ,S,S}(U_1, V_2, V_3, U_4)\\
&+ S_{MZ,S,S}(V_1, U_2, V_3, U_4) + S_{MZ,S,S}(V_1, V_2, U_3, U_4) - S_{MZ,S,S}(V_1, V_2, V_3, U_4)\\
&- S_{MZ,S,S}(V_1, V_2, U_3, V_4) - S_{MZ,S,S}(V_1, U_2, V_3, V_4) - S_{MZ,S,S}(U_1, V_2, V_3, V_4)\\
&+ S_{MZ,S,S}(V_1, V_2, V_3, V_4).
\end{aligned}
$$

The expression of the likelihood for a monozygotic twin and one sibling can be found by setting $U_4 = 0$ and $V_4 = \infty$ ($S_{MZ,S,S}(x_1, x_2, x_3, \infty) = 0$ and $S_{MZ,S,S}(x_1, x_2, x_3, 0) = S_{MZ,S}(x_1, x_2, x_3)$). The likelihood simplifies to

$$
\begin{aligned}
&\mathrm{lik}_{MZ,S}((U_1, V_1), (U_2, V_2), (U_4, V_4))\\
&\quad = S_{MZ,S}(U_1, U_2, U_3) - S_{MZ,S}(U_1, U_2, V_3) - S_{MZ,S}(U_1, V_2, U_3)\\
&\qquad - S_{MZ,S}(V_1, U_2, U_3)\\
&\qquad + S_{MZ,S}(U_1, V_2, V_3) + S_{MZ,S}(V_1, U_2, V_3) + S_{MZ,S}(V_1, V_2, U_3)\\
&\qquad - S_{MZ,S}(V_1, V_2, V_3).
\end{aligned}
$$

For only a monozygotic twin, the likelihood is proportional to (set $U_3 = 0$ and $V_3 = \infty$ in the expression of $\mathrm{lik}_{MZ,S}$):

$$
\begin{aligned}
&\mathrm{lik}_{MZ}((U_1, V_1), (U_2, V_2))\\
&\quad = S_{MZ}(U_1, U_2) - S_{MZ}(U_1, V_2) - S_{MZ}(V_1, U_2) + S_{MZ}(V_1, V_2).
\end{aligned}
$$

The simultaneous survival functions S_{MZ}, $S_{MZ,S}$ and $S_{MZ,S,S}$ can be expressed into the marginal survival function S as described before.

For a couple, a trio or a foursome consisting of a dizygotic twin and zero, one or two siblings similar expressions for the marginal likelihood hold. For the whole sample of n families the likelihood is proportional to the product of all likelihoods of the couples, trios and foursomes, by independence of the families.

The marginal survival functions for male twins, female twins, male non-twins, and female non-twins may be different. However, we assume that they all belong to the family of truncated and shifted exponential distributions: so their cumulative distribution function has the form $x \to \alpha(1 - \exp(-\lambda(x - s)))$ for $x \geq s$, for α, λ and s unknown parameters. This choice of the family was based on the forms of the non parametric maximum likelihood estimators (NPMLE) of the cumulative distribution functions under the (incorrect) assumption of independent survival data of all individuals. Under the independence assumption, the likelihood for the four data-sets (male twins, female twins, male non-twins and for female non-twins) have the form $\prod_{i=1}^{N}(S(U_i) - S(V_i))$ for N the total number of individuals in the data-set.

To maximize the likelihood with respect to all unknown parameters, including the parameters h^2, c^2, b^2 and σ^2 (and possibly also the gender-specific parameters h_m^2, h_f^2, c_m^2, c_f^2, b_m^2 and b_f^2), we used an iterative algorithm. We maximized iteratively with respect to the three parameters of the marginal survival functions of male twins, next to those of the survival function of the female twins, then of the male siblings, of the female siblings and finally to the frailty parameters (h^2, c^2, b^2, σ^2) until the

value of the likelihood did not increase anymore. We performed the maximization with different starting values and different order of maximization of the parameters.

3.2 Testing

To test whether the age at which individuals visit a social worker for the first time is heritable (i.e. whether $h^2 > 0$), whether this age is affected by common environmental factors (i.e. whether $c^2 > 0$) or whether there is a twin-effect (whether $b^2 > 0$), we test the following hypotheses sequentially

$$\begin{aligned} H_0 : h^2 = 0 \quad &\vee \quad H_1 : h^2 > 0, \\ H_0 : c^2 = 0 \quad &\vee \quad H_1 : c^2 > 0, \\ H_0 : b^2 = 0 \quad &\vee \quad H_1 : b^2 > 0. \end{aligned}$$

Note that the three null hypotheses are equivalent to the hypotheses that $\nu = 0$, $\nu_c = 0$ and $\nu_b = 0$, respectively. The hypotheses are tested by a likelihood ratio test. In the numerator of the likelihood ratio statistic the likelihood is maximized over the full parameter space, whereas in the denominator the likelihood is maximized over the restricted space with h^2, c^2 or b^2 equal to zero (depending on the hypothesis that is tested). Since we assume that the survival functions of the male twins, female twins, male non-twins, and female non-twins are parametric distributions (truncated and shifted exponential distributions), the asymptotic distribution of the likelihood ratio test statistic is a 50–50% mixture distribution of a point mass at zero and a chi-square distribution with one degree of freedom (see, e.g., van der Vaart 1998, Chap. 16). For $\alpha < 1/2$ (with α the level of the test), the upper α-quantile of this mixture distribution is the upper 2α-quantile of the chi-squared distribution with one degree of freedom. Parameters that are not significantly bigger than zero are taken equal to zero in the model (the level of the test was taken equal to 0.05).

Tests whether one of the effects is gender specific can be performed similarly (after including the gender specific term into the model).

The last question is whether the cumulative distribution functions for the age at which twins and their non-twin siblings ask for social help, differ. Define F_T^M, F_T^F, F_S^M, and F_S^F as the cumulative distribution functions for male twins, female twins, male siblings and female siblings. The two hypotheses that have to be tested are:

$$\begin{aligned} H_0 : F_T^F \equiv F_S^F \quad &\vee \quad H_1 : F_T^F \not\equiv F_S^F, \\ H_0 : F_T^M \equiv F_S^M \quad &\vee \quad H_1 : F_T^M \not\equiv F_S^M. \end{aligned}$$

Since the four distribution functions are known up to the truncation, shift and intensity parameter, it is actually tested whether the three parameters of the distribution functions for (fe)male twins and (fe)male non-twins equal. Under both null hypotheses the likelihood ratio test statistic has asymptotically a chi-squared distribution with three degrees of freedom. It can also be tested whether the distribution functions of twins

and non-twins differ at a certain age, for instance at age 50. Under the null hypothesis that the distribution functions equal at the particular age, the limit distribution of the likelihood ratio test-statistic is a chi-square distribution with one degree of freedom.

4 Results

4.1 The data

In an ongoing study on health, lifestyle, and personality longitudinal data were collected in a large sample of Dutch twins and their families. The participants were volunteer members of the Netherlands Twin Registry, kept by the department of Biological Psychology at the VU University in Amsterdam (Boomsma 2002; Boomsma et al. 2006). Surveys were mailed to the participants at five different time points between 1991 and 2002. In each of these surveys, questions were asked whether the participants ever had contacted a social worker for non-physical problems. From the data we selected all families with information of at least one twin halve (a monozygotic or a dizygotic twin). Some of the individuals filled in all five questionnaires, others only a few of them. In total, data of 4,499 families is available. If a family consisted of a twin and more than two siblings only two of the siblings were included into the data. Of these 4,499 families 727 contained information on only one individual, 1,988 on two individuals, 1,232 on three, and 551 families contained information on four individuals. Of the 4,499 families, 1,917 consisted of a monozygotic twin and possibly siblings and 2,582 of a dizygotic twin and siblings. In total we have data of 6,289 females and 4,314 males.

The non-parametric maximum likelihood estimators (NPMLEs) of the cumulative distribution functions for the age at which males and females contact a social worker for the first time were computed under the assumption that the data of all males and of all females are independent (see Fig. 1). This was also done for the twins and non-twins separately (see Fig. 2).

There is no information under the age of 10 and hardly any above the age of 65. So, parametric and non-parametric estimates of the cumulative distribution function below the age of 10 and above the age of 65 are not reliable. Estimates of life time risks are, therefore, also not reliable.

The frailty decomposition in the final model contains the terms for additive genetic effects, common environmental effects and twin effects; they were all significant. The p-values of the corresponding three tests equalled 1.18×10^{-6}, 0.0198 and 0.000361, respectively. The extra terms in the frailty for gender specific effects were not included in the final model, because these effects were not significant (the smallest p-value was: 0.294). Furthermore, the null hypothesis that the cumulative distribution functions for female twins and non-twins equal was not rejected (p-value: 0.282). In the final model we assume that the age at which female twins and female non-twins visit a social worker for the first time equal. For the male twins and non-twins the equivalent null hypothesis was rejected (p-value: 0.000264); in the final model different survival functions are modeled for male twins and non-twins. Also the null hypotheses that the male

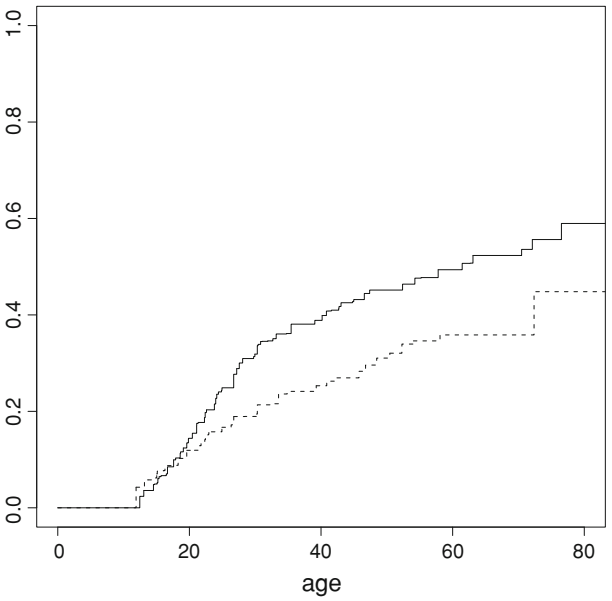

Fig. 1 Non-parametric maximum likelihood estimates of the cumulative distribution function of the age at which females and males contact a social worker. *Continued step-function*: Females. *Dashed step-function*: males

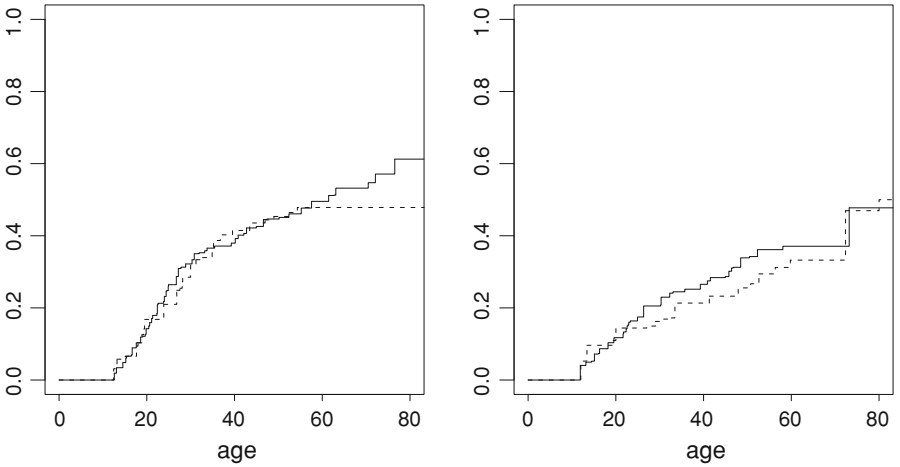

Fig. 2 Non-parametric maximum likelihood estimates of the cumulative distribution function of the age at which females (*left*) and males (*right*) contact a social worker. *Continued curves*: twins, *Dashed curves*: siblings. Number of observations: male twins: 3,141, female twins: 4,948, male non-twins: 1,173, female non-twins: 1,341

distribution functions equalled at the age of 50 or 65 were rejected (p-values: 0.00224 and 0.0285).

The maximum likelihood estimates of the parameters h^2, c^2, b^2 and σ in the final model are given by $\hat{h}^2 = 0.392$, $\hat{c}^2 = 0.101$, $\hat{b}^2 = 0.137$, and $\hat{\sigma} = 3.095$. The parameter estimates of the three cumulative distribution functions are given in Table 1

Table 1 Maximum likelihood estimates of the parameters of the three distribution functions under the assumption that these distribution functions belong to the family of truncated and shifted exponential distributions

	α	shift	λ
Females	0.682	12.21	0.0329
Male twins	0.640	10.13	0.0212
Male non-twins	0.997	2.02	0.0071

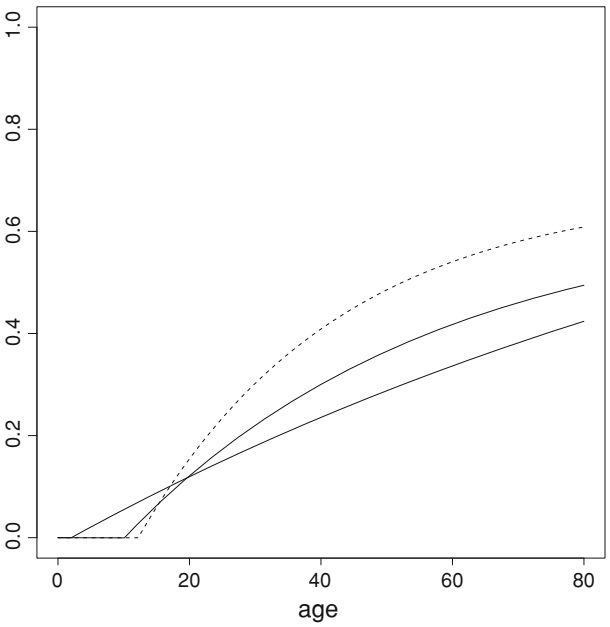

Fig. 3 Maximum likelihood estimates of the cumulative distribution function of the age at which females and males contact a social worker. *Dashed curves*: females, *Continued curves*: male-twins (*upper*), male-siblings (*lower*)

and the estimated curves are shown in Fig. 3 (it is assumed that the survival functions for female twins and female non-twins equal). The estimated distribution function for male non-twins almost starts at the origin, whereas the other two distribution functions are shifted more than 10 years to the right. As mentioned before, the estimated distributions functions are not reliable for young ages, because before the age of 10 no information is available. We compared the parametric and the non-parametric maximum likelihood estimated curves for male non-twins. In all cases the two distribution functions were very similar at the interval from 15 to 50 years. Above the age of 50 the estimated curves diverge slightly, probably because the amount of information is small.

5 Discussion

In this paper we considered an additive gamma frailty model for survival data of twins and siblings. The frailty term was decomposed into four terms: genetic effect, common environmental effect of all twins and non-twins in the family, twin effect and individual specific genetic and environmental effect. The presented model is a nice alternative for variance decomposition models (see e.g., Sham 1998) in case the survival data are censored or truncated. In the variance decomposition model the trait value is decomposed into a linear combination of independent normally distributed random variables. In the model presented in this paper the frailties are decomposed; the frailties are viewed as latent phenotypes.

The major aim in this paper was to test whether twins more or less often ask for social help than their non-twin brothers and sisters. For females we found no significant difference between the cumulative distribution functions. For males a difference was found; before the age of 50 more twins than non-twin brothers visited a social worker. The same result was found at the age of 65.

With the presented model survival data of a twin and zero, one or two siblings is modeled. The model can be easily extended to deal with more than two siblings. Other family members can also be included into the model, but then the frailty decomposition has to be changed to deal with cohort-effects. Moreover, the marginal survival function of other family members, like parents, may be completely different, because they grew up in a different period with possibly a different philosophy concerning social help.

For mathematical convenience we assumed that the frailty variable equals a sum of independent gamma distributed variables with a common scale parameter, so that the frailty itself is gamma distributed again. Only if the frailty has a gamma distribution the multivariate survival function can be written in terms of the marginal distributions. Further research to alternative models for the frailty variable is needed.

We assumed that the survival functions belong to a family of parametric distributions. When we are interested in the frailty parameters, we could, as an alternative, estimate the survival functions by their NPMLEs, insert them into the expression of the likelihood or likelihood ratio statistic and next maximize with respect to the frailty parameters. When maximizing the likelihood with respect to the frailty parameters the estimates of the survival functions are taken fixed. However, in this case the asymptotic distribution of the likelihood ratio test-statistic is unknown and the values of the test-statistic can not be compared with the quantiles of a chi-squared distribution in order to find p-values. Moreover, testing whether the survival functions for twins and non-twins equal is not possible.

In this paper we considered a statistical model with four possibly different survival functions (for male twins, female twins, male non-twins and female non-twins). In general this approach works well for discrete covariates with only a few categories (in our case male/female and twin/non-twin). However, it is not applicable in case of continuous covariates or discrete covariates with many categories. Then a proportional hazards model could be used.

Appendix: derivation of expressions for $S_{MZ,S}$ and $S_{DZ,S}$

For simplicity of notation we derive an expressions of $S_{MZ,S}$ in terms of the marginal survival function S. The derivations of expressions for $S_{DZ,S}$, $S_{MZ,S,S}$ and $S_{DZ,S,S}$ in terms of S are similar.

The trio of frailties (Z_1, Z_2, Z_3) as defined in Sect. 2 has a three dimensional Gamma distribution with marginal distributions equal to the $\Gamma(\nu + \nu_c + \nu_d + \nu_e, \eta)$ distribution with $\eta = \nu + \nu_c + \nu_b + \nu_e$. This marginal distribution is infinitely divisible. Infinity divisible distributions correspond one-to-one with Lévy processes. These are processes $Y = (Y_t : t \geq 0)$ with stationary, independent increments and $Y_0 = 0$. The corresponding infinitely divisible distribution is the distribution of Y_1. Given independent copies Y, $\bar{Y}$ and $\tilde{Y}$, we define the vector with frailty variables (Z_1, Z_2, Z_3) by

$$\begin{aligned} &Y_{\nu/(2\eta)} + \left(Y_{\nu/\eta} - Y_{\nu/(2\eta)}\right) + \left(Y_{(\nu+\nu_c)/\eta} - Y_{\nu/\eta}\right) \\ &\quad + \left(Y_{(\nu+\nu_c+\nu_b)/\eta} - Y_{(\nu+\nu_c)/\eta}\right) + \left(Y_1 - Y_{(\nu+\nu_c+\nu_b)/\eta}\right), \end{aligned} \tag{8}$$

for the first coordinate, Z_1,

$$\begin{aligned} &Y_{\nu/(2\eta)} + \left(Y_{\nu/\eta} - Y_{\nu/(2\eta)}\right) + \left(Y_{(\nu+\nu_c)/\eta} - Y_{\nu/\eta}\right) \\ &\quad + \left(Y_{(\nu+\nu_c+\nu_b)/\eta} - Y_{(\nu+\nu_c)/\eta}\right) + \left(\bar{Y}_1 - \bar{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right), \end{aligned} \tag{9}$$

for the second coordinate, Z_2, and

$$\begin{aligned} &\tilde{Y}_{\nu/(2\eta)} + \left(Y_{\nu/\eta} - Y_{\nu/(2\eta)}\right) + \left(Y_{(\nu+\nu_c)/\eta} - Y_{\nu/\eta}\right) \\ &\quad + \left(\tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta} - \tilde{Y}_{(\nu+\nu_c)/\eta}\right) + \left(\tilde{Y}_1 - \tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right), \end{aligned} \tag{10}$$

for the third coordinate, Z_3. The first and the second term in the decompositions correspond with the additive genetic frailty component A, the third term with common environment component C, the fourth term with the term for twin effect B, and the last term with the non-shared, specific environmental and genetic effects E. The decompositions in the previous displays were constructed so that the correlation of the genetic terms between the two MZ twins halves equals 1 and between two siblings (who do not form an MZ-twin) equals 1/2, the correlation of the common environmental effects between two siblings equals 1, the correlation of the twin effect between two twins equals 1 and 0 otherwise, and the correlation of non-shared, specific environmental and genetic effects between the individuals always equals 0.

The frailty terms Z_1, Z_2 and Z_3 possess marginally the same distribution and the mutual correlations equal

$$\begin{aligned}\rho_{MZ} &= \mathrm{Cor}(Z_1, Z_2) = \frac{\mathrm{Cov}(Z_1, Z_2)}{\mathrm{Var}\ Z_1} = \frac{\mathrm{Var}\ Y_{(\nu+\nu_c+\nu_b)/\eta}}{\mathrm{Var}\ Y_1} \\ &= \frac{\nu+\nu_c+\nu_b}{\eta} = h^2 + c^2 + b^2, \\ \rho_{MZ,S} &= \mathrm{Cor}(Z_1, Z_3) = \frac{\mathrm{Cov}(Z_1, Z_3)}{\mathrm{Var}\ Z_1} = \frac{\mathrm{Var}\ (Y_{(\nu+\nu_c)/\eta} - Y_{\nu/2\eta})}{\mathrm{Var}\ Y_1} \\ &= \frac{\nu/2+\nu_c}{\eta} = \frac{1}{2}h^2 + c^2.\end{aligned}$$

An equal expression holds for $\mathrm{Cor}(Z_2, Z_3)$. In order to obtain non-negative frailties, the distribution of Y_1 and consequently also of $\bar{Y}_1$ and $\tilde{Y}_1$ must be concentrated on $[0, \infty)$.

With the definition of the frailties as just defined, the joint survival function of (T_1, T_2, T_3) is given by

$$\begin{aligned}&S_{MZ,S}(t_1, t_2, t_3) := \mathrm{P}(T_1 > t_1, T_2 > t_2, T_3 > t_3) \\ &= \mathrm{E}\left(\mathrm{e}^{-Z_1\Lambda(t_1)}\mathrm{e}^{-Z_2\Lambda(t_2)}\mathrm{e}^{-Z_3\Lambda(t_3)}\right) \\ &= \mathrm{E}\Big(\mathrm{e}^{Y_{\nu/2\eta}(\Lambda(t_1)+\Lambda(t_2))}\mathrm{e}^{(Y_{(\nu+\nu_c+\nu_b)/\eta}-Y_{(\nu+\nu_c)/\eta})(\Lambda(t_1)+\Lambda(t_2))} \\ &\quad\times \mathrm{e}^{(Y_{(\nu+\nu_c)/\eta}-Y_{\nu/2\eta})(\Lambda(t_1)+\Lambda(t_2)+\Lambda(t_3))}\mathrm{e}^{(Y_1-Y_{(\nu+\nu_c+\nu_b)/\eta})\Lambda(t_1)}\mathrm{e}^{\left(\bar{Y}_1-\bar{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right)\Lambda(t_2)} \\ &\quad\times \mathrm{e}^{\tilde{Y}_{\nu/2\eta}\Lambda(t_3)}\mathrm{e}^{\left(\tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta}-\tilde{Y}_{(\nu+\nu_c)/\eta}\right)\Lambda(t_3)}\mathrm{e}^{\left(\bar{Y}_1-\tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right)\Lambda(t_3)}\Big) \\ &= \mathrm{E}\Big(\mathrm{e}^{(Y_{\nu/2\eta}+(Y_{(\nu+\nu_c+\nu_b)/\eta}-Y_{(\nu+\nu_c)/\eta}))(\Lambda(t_1)+\Lambda(t_2))}\mathrm{e}^{(Y_{(\nu+\nu_c)/\eta}-Y_{\nu/2\eta})(\Lambda(t_1)+\Lambda(t_2)+\Lambda(t_3))} \\ &\quad\times \mathrm{e}^{(Y_1-Y_{(\nu+\nu_c+\nu_b)/\eta})\Lambda(t_1)}\mathrm{e}^{\left(\bar{Y}_1-\bar{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right)\Lambda(t_2)} \\ &\quad\times \mathrm{e}^{\left(\tilde{Y}_{\nu/2\eta}+\left(\tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta}-\tilde{Y}_{(\nu+\nu_c)/\eta}\right)+\left(\bar{Y}_1-\tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right)\right)\Lambda(t_3)}\Big) \\ &= \mathrm{E}\left(\mathrm{e}^{(Y_{\nu/2\eta}+(Y_{(\nu+\nu_c+\nu_b)/\eta}-Y_{(\nu+\nu_c)/\eta}))(\Lambda(t_1)+\Lambda(t_2))}\right) \\ &\quad\times \mathrm{E}\left(\mathrm{e}^{(Y_{(\nu+\nu_c)/\eta}-Y_{\nu/2\eta})(\Lambda(t_1)+\Lambda(t_2)+\Lambda(t_3))}\right) \\ &\quad\times \mathrm{E}\left(\mathrm{e}^{(Y_1-Y_{(\nu+\nu_c+\nu_b)/\eta})\Lambda(t_1)}\right)\mathrm{E}\left(\mathrm{e}^{\left(\bar{Y}_1-\bar{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right)\Lambda(t_2)}\right) \\ &\quad\times \mathrm{E}\left(\mathrm{e}^{\left(\tilde{Y}_{\nu/2\eta}+\left(\tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta}-\tilde{Y}_{(\nu+\nu_c)/\eta}\right)+\left(\bar{Y}_1-\tilde{Y}_{(\nu+\nu_c+\nu_b)/\eta}\right)\right)\Lambda(t_3)}\right) \\ &= \psi(\Lambda(t_1)+\Lambda(t_2))^{(\nu/2+\nu_b)/\eta}\psi(\Lambda(t_1)+\Lambda(t_2)+\Lambda(t_3))^{(\nu/2+\nu_c)/\eta} \\ &\quad\times \psi(\Lambda(t_1))^{\nu_e/\eta}\psi(\Lambda(t_2))^{\nu_e/\eta}\psi(\Lambda(t_3))^{(\nu/2+\nu_b+\nu_e)/\eta},\end{aligned}$$

for $\psi(u) = \mathrm{E}e^{-uY_1}$, the Laplace transform of Y_1. In the last equality we used that $\nu_e = \eta - \nu - \nu_c - \nu_b$ and the identity $\mathrm{E}e^{-uY_t} = (\mathrm{E}e^{-uY_1})^t$, which follows from the independence and the stationarity of the increments. The Laplace transform of the frailties equals $\psi(u) = (1 + u/\eta)^{-\eta}$. The corresponding joint survival function is, then, given by

$$
\begin{aligned}
&S_{MZ,S}(t_1,t_2,t_3) = \mathrm{P}(T_1 > t_1, T_2 > t_2, T_3 > t_3) \\
&= \left(\frac{1}{1+\Lambda(t_1)/\eta+\Lambda(t_2)/\eta}\right)^{\nu/2+\nu_b} \left(\frac{1}{1+\Lambda(t_1)/\eta+\Lambda(t_2)/\eta+\Lambda(t_3)/\eta}\right)^{\nu/2+\nu_c} \\
&\quad\times \left(\frac{1}{1+\Lambda(t_1)/\eta}\right)^{\nu_e} \left(\frac{1}{1+\Lambda(t_2)/\eta}\right)^{\nu_e} \left(\frac{1}{1+\Lambda(t_3)/\eta}\right)^{\nu/2+\nu_b+\nu_e}.
\end{aligned}
$$

Setting $t_2 = t_3 = 0$ shows that the marginal conditional survival function is given by

$$
S(t) = \mathrm{P}(T > t) = S_{MZ,S}(t,0,0) = \left(\frac{1}{1+\Lambda(t)/\eta}\right)^{\eta},
$$

is dependent of the parameters ν, ν_c, ν_b and ν_e only via the sum $\eta = \nu + \nu_c + \nu_b + \nu_e$. Solving $\Lambda(t)$ from the equation in the previous display, yields

$$
\Lambda(t) = \eta\left(S(t)^{-1/\eta} - 1\right).
$$

Substituting this equality into the expression of $S_{MZ,S}$ and taking $\sigma^2 = 1/\eta$ yields the expression in (4).

An expression for $S_{DZ,S}(t_1,t_2,t_3)$ can be found in a similar way. We define the vector with frailty variables (Z_1, Z_2, Z_3) by, respectively, (8), (9), and (10), where the first two terms (the genetic term) are replaced by, respectively

$$
\begin{aligned}
&Y_{\nu/(4\eta)} + \left(Y_{2\nu/(4\eta)} - Y_{\nu/(4\eta)}\right) + \left(Y_{3\nu/(4\eta)} - Y_{2\nu/(4\eta)}\right) + \left(Y_{\nu/\eta} - Y_{3\nu/(4\eta)}\right), \\
&Y_{\nu/(4\eta)} + \left(Y_{2\nu/(4\eta)} - Y_{\nu/(4\eta)}\right) + \left(\bar{Y}_{3\nu/(4\eta)} - \bar{Y}_{2\nu/(4\eta)}\right) + \left(\bar{Y}_{\nu/\eta} - \bar{Y}_{3\nu/(4\eta)}\right), \\
&Y_{\nu/(4\eta)} + \left(\bar{Y}_{2\nu/(4\eta)} - \bar{Y}_{\nu/(4\eta)}\right) + \left(\bar{Y}_{3\nu/(4\eta)} - \bar{Y}_{2\nu/(4\eta)}\right) + \left(Y_{\nu/\eta} - Y_{3\nu/(4\eta)}\right).
\end{aligned}
$$

This time the decomposition of the genetic term is chosen so that the correlation of this term between every two siblings equals 1/2 (two siblings who do not form a MZ-twin share on average half of their genes). Then,

$$
\begin{aligned}
\rho_{DZ} &= \mathrm{Cor}(Z_1,Z_2) = \frac{\mathrm{Cov}(Z_1,Z_2)}{\mathrm{Var}\,Z_1} = \frac{\mathrm{Var}\,Y_{\nu/(2\eta)} + \left(Y_{(\nu+\nu_c+\nu_b)/\eta} - Y_{\nu/\eta}\right)}{\mathrm{Var}\,Y_1} \\
&= \frac{\nu/2+\nu_c+\nu_b}{\eta} = \frac{1}{2}h^2 + c^2 + b^2, \\
\rho_{DZ,S} &= \mathrm{Cor}(Z_1,Z_3) = \frac{\mathrm{Cov}(Z_1,Z_3)}{\mathrm{Var}\,Z_1} = \frac{\mathrm{Var}\,Y_{\nu/(4\eta)} + \left(Y_{(\nu+\nu_c)/\eta} - Y_{3\nu/(4\eta)}\right)}{\mathrm{Var}\,Y_1} \\
&= \frac{\nu/2+\nu_c}{\eta} = \frac{1}{2}h^2 + c^2.
\end{aligned}
$$

Similar computations show that $\mathrm{Cor}(Z_2,Z_3) = (\nu/2+\nu_c)/\eta$. The remaining derivation follows the same line as for $S_{MZ,S}$.

Acknowledgments The project was financially supported by NDNS+. We thank Piet Groeneboom for using his computer program for computing the NPMLE for a survival curve based on interval censored survival data. We also thank Sandjai Bhulai for his help writing the computer program for maximizing the likelihood function. Last, but not least, we thank the twins and their family members for filling in the questionnaires.

References

Andersen PK, Borgan O, Gill RD, Keiding N (1992) Statistical models based on counting processes. Springer, New York

Boomsma DI (2002) Netherlands twin register: a focus on longitudinal research. Twin Res 5:401–406

Boomsma DI, Geus EJC, de Vink JM, Stubbe JH, Disteland MA, Hottenga JJ (2006) Netherlands twin register: from twins to twin families. Twin Res Hum Genet 9:849–857

Iachine I (2001) The use of twin and family survival data in the population studies of aging: statistical methods based on multivariate survival models. Dissertation, Department of Statistics and Demography, University of Southern Denmark

Jonker MA, Bhulai S, Ligthart RSL, Posthuma D, Boomsma DI, van der Vaart AW (2009) Gamma frailty model for linkage analysis with application to interval censored migraine data. Biostatistics 10:187–200

Sham PC (1998) Human genetics. Arnold Publishers, London

Vaupel JW, Manton KG, Stallard E (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16(3):439–454

van der Vaart AW (1998) Asymptotic Statistics. Cambridge University Press, Cambridge

Yashin AI, Iachine IA (1999) Dependent hazards in multivariate survival problems. J Multivar Anal 71:241–261

Yashin AI, Vaupel JW, Iachine IA (1995) Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. Math Popul Stud 5(2):145–159

Yashin AI, Begun AZ, Iachine IA (1999) Genetic factors in susceptibility to death: a comparitive analysis of bivariate survival models. J Epidemiol Biostat 4(1):53–60